Preparing a Library of Compounds for Use by Glide or Phase

This document contains instructions on how to prepare a virtual screening library of compounds for use by Glide™ or by Phase™. While this procedure can be varied, the instructions given below provide a good rule-of-thumb.

The steps in this process are as follows:

1. Obtain 2D SD files
2. Prepare structures
3. Filter structures (optional)
4. Perform product-specific processing

1. Obtain 2D SD files

If you do not already have a library of compounds in 2D SD format, you can find links to sources of compounds at the HTScreening.net web site, http://www.htscreening.net/home, or the ZINC web site, http://blaster.docking.org/zinc.

If the library contains less than 50,000 compounds, you can continue with the procedure. If the library is larger, you must split it into separate files, each containing about 50,000 compounds, and process each file. You can split the file using the `sdsubset` utility:

`$SCHRODINGER/utilities/sdsubset -n` *range* *full*`.sdf >` *subset*`.sdf`

Here, *range* specifies the first and last structure to extract, separated by a colon, e.g. `1:50000`.

2. Prepare structures

The structures in your 2D SD files must be converted into 3D all-atom structures, using LigPrep. The LigPrep actions performed in this part of the preparation are:

- Retain the specified chiralities
- Generate a maximum of 4 stereoisomers
- Remove cofactors (desalt)
- Neutralize charged acidic or basic groups
- Generate only the lowest-energy ring conformation
- For Glide, generate tautomers

The output should be generated in Maestro format. The command to run this job is as follows.

Glide: `$SCHRODINGER/ligprep -i 1 -r 1 -s 4 -isd` *infile*`.sdf -omae` *outfile*`.mae`

Phase: `$SCHRODINGER/ligprep -i 1 -r 1 -s 4 -nt -isd` *infile*`.sdf -omae` *outfile*`.mae`

Here, *infile*`.sdf` is the name of the 2D SD file, and *outfile*`.mae` is the name of the Maestro output file.

3. Filter structures (optional)

After preparing the structures, you might want to filter the structures, based on their ADME or toxicology properties, for example. Below are examples of the use of QikProp to calculate ADME properties and filtering based on these properties.

To calculate the QikProp properties, run the following command on the files generated in the last step:

`$SCHRODINGER/qikprop -nosa` *filename*`.mae`

The output structure file is named *filename*`-out.mae`, and contains all the properties and descriptors generated by QikProp.

Two filtering schemes are provided below, which use the utility `propfilter`. You can also create your own scheme. The input file for `propfilter` is the output file from QikProp, *filename*`-out.mae`.

- **Lipinski filter**

  This filter passes compounds with molecular weight ≤ 500, octanol/water partition coefficient log $P_{o/w}$ ≤ 5, 5 or fewer hydrogen bond donors, and 10 or fewer hydrogen bond acceptors:

  ```
  $SCHRODINGER/utilities/propfilter -e "r_qp_mol_MW <= 500" -e "r_qp_QPlogPo/w <= 5"
      -e "r_qp_donorHB <= 5" -e "r_qp_hb_acceptHB <= 10" -o filter.mae filename-out.mae
  ```

- **Coarse filter to remove high risk compounds**

  This filter passes compounds with molecular weight in the range 150–650, octanol/water partition coefficient in the range –3 to 6.5, aqueous solubility logS ≥ –7, and polar surface area FISA ≤ 175:

  ```
  $SCHRODINGER/utilities/propfilter -e "r_qp_mol_MW >= 150" -e "r_qp_mol_MW <= 650"
      -e "r_qp_QPlogPo/w >= -3" -e "r_qp_QPlogPo/w <= 6.5" -e "r_qp_QPlogS >= -7"
      -e "r_qp_FISA <= 175" -o filter.mae outfile-out.mae
  ```

Phase Databases

To search a database using Phase, conformers and pharmacophore sites must be generated for each structure. If you intend to use the same set of pharmacophore features for many searches, creating and storing the conformers and sites in the Phase database minimizes the search time. The following procedure is one way in which you can create and store conformers and sites. For details on the commands and options, see Chapter 13 of the *Phase User Manual*. A summary is available in the quick reference guide, *Phase Command-Line Database Management and Searching*.

1. Create a database with the output file from the previous step, which is here called *filename*.mae.

   ```
   $SCHRODINGER/utilities/phasedb_manage -new mae -db dbname -confs false -mae filename.mae
   ```

   The database is created in the current directory, and the files for each structure are stored in a subdirectory, *dbname*_ligands. If you have multiple sets of structures, you can add them to the database as follows:

   ```
   $SCHRODINGER/utilities/phasedb_manage -add -db dbname -confs false -mae filename2.mae
   ```

2. Generate conformers and pharmacophore sites for the structures in the database:

   ```
   $SCHRODINGER/utilities/phasedb_confsites -db dbname -JOB jobname -confs all
   ```

   This job generates the conformers and sites with the default options.

For more details on database creation, see the tutorial provided with the Phase installation:

```
$SCHRODINGER/phase-vversion/tutorial/db_tutorial.pdf
```

Glide Ligand Libraries

For Glide, all that is required after running the preparation procedure is to rerun LigPrep to prepare the structures in the appropriate ionization state. The LigPrep actions performed in this part of the preparation are:

- Determine chiralities from the 3D structure
- Do not generate stereoisomers
- Do not remove cofactors (desalt)
- Generat possible ionization states at pH 7.0 2.0
- Generate only the lowest-energy ring conformation
- Do not generate tautomers

The command to run this job is as follows.

```
$SCHRODINGER/ligprep -g -i 2 -r 1 -s 1 -nd -nt -imae infile.mae -omae outfile.mae
```